

基于大数据的教育决策支持研究*

王 博¹ 魏顺平²【通讯作者】



(1. 南京工业职业技术学院 发展规划办公室 江苏南京 210023;

2. 国家开放大学 信息化部(工程中心), 北京 100039)

摘要: 文章以推动大数据应用为出发点, 选取基于大数据的教育决策支持为应用切入点, 系统梳理了决策及决策支持的相关概念、大数据对于决策的种种影响、可用的技术及存在的不足; 进而提供了基于大数据的决策支持模式, 并对该模式的特点包括全面化的数据采集、高效深度的数据处理、可视化结果呈现等进行了阐述; 最后以远程教育领域人才引进决策支持为例, 全面展现了基于大数据的决策支持过程。

关键词: 大数据; 教育决策支持; 模式; 案例

【中图分类号】G40-057 【文献标识码】A 【论文编号】1009—8097(2016)04—0005—07 【DOI】10.3969/j.issn.1009-8097.2016.04.001

引言

2015年10月29日, 中共中央十八届五中全会顺利闭幕。会议通过了“十三五”计划并出台了一系列重大政策, 启动“互联网+”行动计划, 实施国家大数据战略。2015年9月, 国务院印发《促进大数据发展行动纲要》(以下简称《纲要》), 系统部署大数据发展工作。在《促进大数据发展行动纲要》中, 具体到教育领域, 提出“教育文化大数据工程”。

要充分发挥教育大数据的价值, 关键在于应用, 其中为教育决策提供支持服务、构建科学规范的教育治理体系、推动教育现代化, 无疑是当前紧迫且必要的应用领域。《国家中长期教育改革和发展规划纲要(2010-2020年)》指出, 要“加强教育宏观政策和发展战略研究, 提高教育决策科学化水平”。袁贵仁同志在2014年全国教育工作会上的讲话中, 强调“完善科学规范的教育治理体系, 形成高水平的教育治理能力”、“通过深化教育领域综合改革, 实现教育事业科学发展”。在2016年全国教育工作会议上, 袁贵仁同志又做了“以新的发展理念为引领, 全面提高教育质量, 加快推进教育现代化”为题的讲话。在这些文件中, “现代化”、“科学”、“治理”等字词频繁出现。在发展现代而科学的教育治理体系与治理能力方面, 大数据的应用将有望为教育发展注入新的活力。随着信息技术特别是互联网技术的迅速发展, 基于大数据的政府治理与决策应运而生。在教育领域, 随着大数据发展而带来的教育政策研究与决策的“以证据为本”之趋势亦渐明显^[1]。将大数据引入教育决策, 将有助于其科学水平的提高。

一 大数据给决策带来的影响分析

大数据是继云计算、物联网之后又一新兴的信息技术。李国杰院士等^[2]认为, 大数据是无法在可容忍的时间内用传统IT技术和硬件工具对其进行感知、获取、管理、处理和服务的数据集合。数据量巨大(Volume)、数据新增速度加快(Velocity)、数据来源多样化(Variety)、数据价值密度低(Value)被视为大数据的四个核心特征^[3]。针对大数据的全新特征, 人们在分析数据和信息时要做出相应的转变——从随机抽样向采集全部样本的转变, 从追求精确向掌握大体方向的转变, 从寻找因果关系向寻找相关关系的转变^[4]。

我们知道, 计算机时代产生了计算方式的变革, 互联网时代产生了信息传播方式的革命。那么, 大数据时代也将会带来一个决策方式的革命。《大数据主义》一书提到: “技术革命将改变人类决策方式。人们将更加依赖于数据和分析, 而不是直觉和经验; 同样, 它还将改变领导力和管理的本质”^[5]。维克多·迈尔·舍恩伯格^[6]认为: “大数据开启了一次重大的时代转型。就像望远镜让我们能感受宇宙, 显微镜能让我们观察微生物一样, 大数据正在改变我们的生活以及理解世界的方式。”大数据将改变人类的教育方式、人类的学习方式、教育信息化的研究范式, 并在一定程度上改善人类的思维^[7]。

决策是指决策者为了达到一定的行为目的, 根据决策环境的变化所做出的一些决定^[8]。决策本质上也是人类思维活动, 对思维的影响意味着对决策的影响。李德毅院士在一次演讲中指出, “脑认知的本质等同于大数据认知”, 足见大数据对思维的影响。我们知道, 思维是人脑对客观事物的本质属性和事物之间内在联系的规律性所做出的概括与间接的反映^[9]。思维的组成要素有四个, 即思维加工材料、思维加工方式、思维加工缓存区和思维加工机制。不难理解, 有了大数据及其配套的数据仓库、云计算、数据挖掘等支撑技术, 人类的思维加工材料可以是整个数据仓库, 其内容将变得极大丰富; 思维加工缓存区可以是整个云计算服务器, 其吞吐能力将大幅提升; 思维加工方式可以应用各种数据挖掘算法, 其处理方式将变得更加丰富多样。以逻辑思维为例, 由于大数据提供了丰富的实证材料使逻辑思维的加工过程可能被缩短了, 或者使逻辑思维的有效性进一步提升了(得出的结论更让人信服了), 也就是能使逻辑思维的效率和质量都得到比较显著的提高或改进^[10]。

在宏观决策层面, 大数据可以发挥诸多价值。以教育政策决策为例, 基于大数据的政策决策有三方面的优势: 首先, 通过大数据可以将微观层面的政治对象呈现出来, 清晰描绘出原本模糊的教育活动, 如此一来, 政策问题得到更好描述; 其次, 大数据实时变化的特点可以使决策者在短时间内获得政策反馈甚至获得实时反馈; 第三, 大数据可以对未来进行预测, 使决策者具有更为开阔的视野。在大数据的时代, 视野已经成为宏观控制的精髓, 而不是力度。大数据是一种动态宏观视野, 能够超越个体与局部的相对静态视野, 更容易发现问题所在、可能弱点和盲区^[11]。

大数据正在颠覆传统的、线性的、自上而下的精英决策模型, 逐步形成非线性的、面向不确定性的、自下而上的决策方式^[12]。这种决策模式遵循“‘万事万物’量化为数据 数据转变为信息 信息转变为知识 知识涌现出智慧”的逻辑, 通常称之为数据化决策。

二 基于大数据的教育决策支持模型

目前, 大数据还无法替代人类完成决策, 更多是提供决策支持。基于统计数据的教育决策支持服务平台, 可以提供教育宏观决策服务, 通过对历史统计数据进行分析, 形成对我国教育发展状况各方面的趋势分析, 给国家制订长远规划提供数据理论依据; 可以提供教育动态监管、预警服务根据教育大数据实时变化情况, 多平台、多时相、多波段和多源数据实时掌控教育动态, 为各种教育专项工程提供全程监管、预警服务; 可以提供教育个体综合评价、教育管理、教学质量评价服务, 通过教育大数据挖掘产生的知识与信息, 传递给知识库管理系统, 使系统智能化、知识化, 实现对教育规律、决策规律以及模型、方法、数据等方面知识的存储和管理, 进而对教育个体、教育管理、教学质量进行评价, 促进教育综合改革的进一步深化。

本文关注的基于大数据的决策支持,主要体现在学校管理工作方面的决策支持。一个一般化的决策支持过程模型符合数据挖掘的一般过程,由“分析工作流”、“工具与算法”、“数据与信息流”三要素构成,如图1所示。在“分析工作流”方面的关键环节有教师绩效评价、人才引进决策、招生决策、就业预测、职业规划、辍学分析、毕业生追踪、课程设置决策等;所用到的“工具与算法”包括统计分析与可视化、相关分析、关联规则、决策树等一系列数据挖掘方法;所用到的“数据与信息”将主要有课堂教学与在线教学数据、教务管理数据、校园生活数据、教职员工数据、后勤管理数据等。

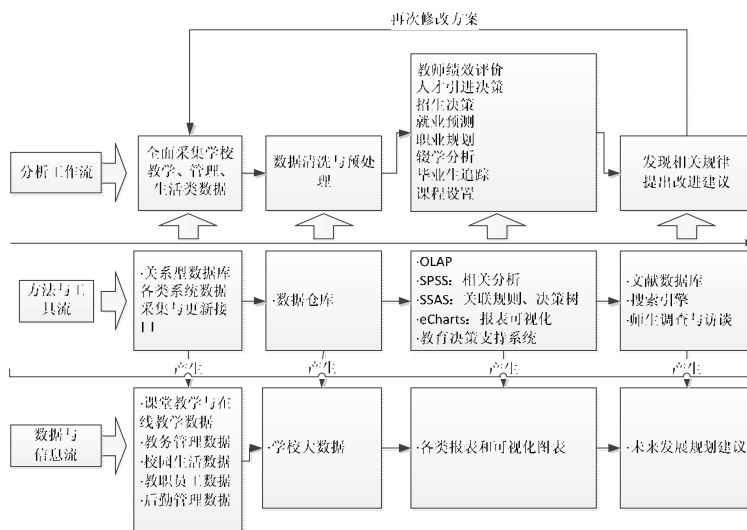


图1 基于大数据的教育决策支持过程模型

这一决策支持模式将体现出“全面化的数据采集、高效深度的数据分析、可视化的结果呈现”等全新特点。

1 全面化的数据采集

教育大数据指的是与教育教学要素相关的以及教育教学过程中产生的全量超大规模、多源异构、实时变化的数据。教育大数据有四大来源：一是在教学活动过程中直接产生的数据，如课堂教学、考试测评、网络互动等；二是在教育管理活动中采集到的数据，如学生的家庭信息、学生的健康体检信息、教职工基础信息、学校基本信息、财物信息、设备资产信息等；三是科学研究活动中采集到的数据，如论文发表、科研设备运行、科研材料采购与消耗等记录信息；四是在校园生活中产生的数据，如餐饮消费、上机上网、复印资料、健身洗浴等记录信息。

2 高效深度的数据分析

教育领域中的数据也逐渐呈现出大数据的“4V”特征。针对数据量巨大（Volume）的特点，可以采用分布式存储和计算方法，相应的工具有Hadoop、Spark、数据仓库以及各类商用大数据服务平台如阿里云、亚马逊AWS等；针对结构化数据、半结构化数据和非结构化数据并存（Variety）的特点，可以采用词语切分、信息抽取等方法，相应的工具有各类自然语言处理工具如大数据搜索与挖掘开发平台；针对数据价值密度低（Value）的特点，可以采用聚类、关联规则、决策树等数据挖掘方法，相应的工具有SQL Server Analysis Service（SSAS）、Weka、SPSS

等；针对数据的产生与处理加速（Velocity）的特点，可以采用信息自动抓取方法，相应的工具有各类网络爬虫软件如火车头采集器。

3 可视化的结果呈现

将分析结果以可视化的形式呈现给决策者，对于决策者理解信息并快速做出决策至关重要。可视化（Visualization）是利用计算机图形学和图像处理技术，将数据转换成图形或图像在屏幕上显示出来，并进行交互处理的理论、方法和技术^[12]。常见的可视化形式有基于坐标的图表、关系图、地理信息图、文字云图、仪表盘等，常用的可视化工具有 Excel、百度 ECharts、UCINET Netdraw、Worditout、R 语言等。

三 远程教育领域案例分享

前文已经提及，人才引进决策是基于大数据的教育决策支持中的一种典型应用。下面以国家开放大学教育学人才引进决策为例，来展现教育大数据的教育决策支持的一般过程，并将在全量数据采集、深度挖掘处理等方面体现大数据的特点。

国家开放大学是教育部直属的，以现代信息技术为支撑，主要面向成人开展远程开放教育的新型高等学校。在师资队伍建设方面，遵循远程教育教学规律，适应国家开放大学建设发展需要，以“专兼结合、以兼为主、动态开放”为原则，建设业务精湛、结构合理、特色鲜明的师资队伍，包括教学、科研、管理、远程学习支持和技术队伍。其中一项重要的工作，就是有重点、有计划地聘请一批国内高校的著名专家教授担任教材主编、课程主讲。那么，聘请哪些专家学者就成为了摆在领导者面前的一项重要决策。

以教育学科为例，国家开放大学应该在全国范围内聘请著名专家教授担任教材主编、课程主讲。学校有必要从大数据的视角来获取全国教育学领域专家学者的相关数据，并分门别类，方便选择、聘任。本案例将选择教育学学科为例，采用标准化（h_iR）指数作为主要指标来对中国大陆地区教育学领域的专家学者（这里“中国大陆地区”用于限定专家学者的学术活动范围，而不是限定专家学者的所在研究机构）进行科研绩效评价，以发现我国教育学领域的主要专家学者，以及这些专家学者在教育学二级学科的分布情况。专家学者都要发表期刊论文，可以基于期刊论文开展科研绩效来衡量专家学者的水平，从而发现著名专家学者。当然，仅仅基于期刊论文相关数据来评价某个领域的专家学者还是有失全面的，因为期刊论文仅仅是专家学者的一种科研产出，本案例的主要价值还是展示大数据在决策支持方面的应用。

1 全面数据采集

对于教育学领域的专家学者，首先要找全，其次要找准。“找全”就要求开展全面化的数据采集，力争获取近 20 年中国大陆地区发表的所有教育学相关期刊论文。本研究按照 1994～2013 年教育学领域重要学术期刊列表（基于历年 CSSCI 教育学科来源期刊整理得来，累计有 46 种期刊），从中国知网期刊论文数据库（CNKI 几乎收录了所有中国大陆地区正式出版的期刊论文）逐年下载对应期刊所刊发的、有被引用记录的论文及其被引用频次（下载时间截止到 2015 年 3 月 1 日），得到的论文共计 114362 篇。

2 数据分析与可视化呈现

（1）评选中国大陆地区教育学领域最具影响力学者

基于前文获得的约 11 万篇研究论文信息，通过拆分论文信息的作者字段，计算每位作者的

决策者做出选择。

表1 最具影响力学者的二级学科分布

聚类结果	学者数	作者列表（以离聚类中心距离为序）
第1类： 高等教育学	29	别敦荣、刘献君、龚放、潘懋元、周远清、王战军、眭依凡、杨叔子、钟秉林、丁小浩、刘海峰、邬大光、岳昌君、张楚廷、顾建民、李曼丽、林健、闵维方、沈红、杨东平、张应强、周济、张民选、阎光才、胡建华、张建伟、黄崑、刘宝存、汪霞
第2类： 教育技术学	21	黄荣怀、顾小清、余胜泉、丁兴富、何克抗、王陆、祝智庭、黎加厚、李克东、南国农、谢幼如、张际平、钟志贤、桑新民、杨开城、赵建华、李芒、李艺、张尧学、高文、陈丽
第3类： 比较教育学	14	顾明远、徐辉、谢安邦、肖川、赵中建、黄甫全、冯建军、袁维新、刘儒德、王鉴、项贤明、朱旭东、崔允漷、钟启泉
第4类： 综合类	36	庞丽娟、辛涛、褚宏启、扈中平、姜勇、申继亮、吴永军、郭元祥、李政涛、卢乃桂、石中英、檀传宝、吴康宁、杨启亮、余文森、朱小蔓、陈向明、孟万金、田慧生、尹弘飏、俞国良、张华、丁邦平、傅维利、霍力岩、靳玉乐、林崇德、宁虹、王坦、吴刚平、郑金洲、叶澜、鲁洁、马云鹏、郭思乐、许明

四 研究结论

十八届五中全会以来，国家启动实施大数据战略，教育领域将相应实施教育文化大数据工程。各级各类教育管理者也应顺应时代发展，充分应用好教育大数据，提升教育决策手段。本文以推动大数据应用为出发点，选取基于大数据的教育决策支持为应用切入点，开展了以下工作：

首先，本文通过梳理相关文献，可知大数据已经对人类决策行为产生了种种影响。大数据可以改变思维的多个要素包括思维加工材料、思维加工缓存区、思维加工方式，还可以改变决策的基本过程。大数据正在颠覆传统的、线性的、自上而下的精英决策模型，逐步形成非线性的、面向不确定性的、自下而上的决策方式。

其次，本文根据现实需要，重点面向学校管理工作方面的决策支持，提出了一个一般化的决策支持过程模型符合数据挖掘的一般过程，由“分析 workflow”、“工具与算法”以及“数据与信息”等三要素构成。这一决策支持模式体现了“全面化的数据采集、高效深度的数据分析、可视化的结果呈现”等全新特点。

最后，本文以远程教育领域人才引进决策支持为例，全面展现了基于大数据的决策支持过程。该案例在教育学研究人员刊发论文全样本数据采集、研究人员聚类等方面充分体现了大数据的特点，并且做到了决策支持。

参考文献

- [1]陈霜叶,孟浏今,张海燕.大数据时代的教育政策证据:以证据为本理念对中国教育治理现代化与决策科学化的启示[J].全球教育展望,2014,(2):121-128.

- [2]李国杰,程学旗.大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J].中国科学院院刊,2012,(6):647-657.
- [3]Mcafee A, Brynjolfsson E. Big data: The management revolution[J]. Boston: Harvard Business Review, 2012,(10):3-9.
- [4][6]维克多·迈尔·舍恩伯格,肯尼恩·库克著.盛扬燕,周涛译.大数据时代——生活、工作与思维的大变革[M].杭州:浙江人民出版社,2013:17-19.
- [5]史蒂夫·洛尔著,胡小锐,朱胜超译.大数据主义[M].北京:中信出版集团,2015:1.
- [7][10]何克抗.大数据面面观[J].电化教育研究,2014,(10):8-16、22.
- [8]柳少军.决策支持理论与方法研究[J].北方交通大学学报,1996,(6):11-16.
- [9]何克抗.创造性思维理论——DC模型的建构与论证[M].北京:北京师范大学出版社,2000:11.
- [11]陈霜叶,孟浏今,张海燕.大数据时代的教育政策证据:以证据为本理念对中国教育治理现代化与决策科学化的启示[J].全球教育展望,2014,(2):121-128.
- [12]陈龙,程开明.大数据时代的决策:数据分析抑或直觉经验[J].中国统计,2014,(9):20-22.
- [13]百度百科.可视化[OL]. <http://baike.baidu.com/link?url=p8N06rCst-ukaCrMh9Q5Q8drSqPCSMlfFdM0lV1VRVEfiy7EgblP4LVm4Hs9m9kXON_eNrJu-8itEabXbnzJPK>

A Study on Educational Decision Support based on Big Data

WANG Bo¹ WEI Shun-ping²(Corresponding Author)

(1. Nanjing Vocational Institute of Industry Technology, Nanjing, Jiangsu, China 210023;

2. R&D Center for E-learning, the Open University of China, Beijing, China 100039)

Abstract: In this paper, the authors selected the promotion of big data application as starting point and focused on educational decision supporting based on big data. An educational decision supporting model based on big data was established by systematically sorting the relevant concepts of decision and decision support, analyzing the influence of big data on decision and the shortages of available technologies. The three characteristics of this model, including comprehensive data collection, efficient and deep data processing and visual representation of result, were expounded. In the end, a case about decision support of talent introduction in distance education field was given to display the decision support process based on big data.

Keywords: big data; educational decision support; model; case

*基金项目:本文为北京市教育科学“十二五”规划2015年度重点课题“基于教育大数据的大规模私有型在线课程教学绩效评估系统及其应用研究”(项目编号:AJA1523)的阶段性研究成果。

作者简介:王博,助理研究员,硕士,研究方向为教育信息化,邮箱为 wangbo756@163.com。

收稿日期:2016年2月20日

编辑:小西